Purpose Restrictions on Information Use

Michael Carl Tschantz, Anupam Datta, and Jeannette M. Wing

June 3, 2013

CMU-CyLab-13-005

CyLab
Carnegie Mellon University
Pittsburgh, PA 15213

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
03 JUN 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Purpose Restrictions on Information Use	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,CyLab,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Privacy policies in sectors as diverse as Web services, finance and healthcare often place restrictions on the purposes for which a governed entity may use personal information. Thus, automated methods for enforcing privacy policies require a semantics of purpose restrictions to determine whether a governed agent used information for a purpose. We provide such a semantics using a formalism based on planning. We model planning using Partially Observable Markov Decision Processes (POMDPs), which supports an explicit model of information. We argue that information use is for a purpose if and only if the information is used while planning to optimize the satisfaction of that purpose under the POMDP model. We determine information use by simulating ignorance of the information prohibited by the purpose restriction, which we relate to noninterference. We use this semantics to develop a sound audit algorithm to automate the enforcement of purpose restrictions.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**17**	

Purpose Restrictions on Information Use*

Michael Carl Tschantz
mct@berkeley.edu
Univ. of California, Berkeley†

Anupam Datta
danupam@cmu.edu
Carnegie Mellon University

Jeannette M. Wing
wing@microsoft.com
Microsoft Research†

June 3, 2013

Abstract

Privacy policies in sectors as diverse as Web services, finance and healthcare often place restrictions on the purposes for which a governed entity may use personal information. Thus, automated methods for enforcing privacy policies require a semantics of *purpose restrictions* to determine whether a governed agent *used information* for a purpose. We provide such a semantics using a formalism based on planning. We model planning using Partially Observable Markov Decision Processes (POMDPs), which supports an explicit model of information. We argue that information use is for a purpose if and only if the information is used while planning to optimize the satisfaction of that purpose under the POMDP model. We determine information use by simulating ignorance of the information prohibited by the purpose restriction, which we relate to noninterference. We use this semantics to develop a sound audit algorithm to automate the enforcement of purpose restrictions.

1 Introduction

Purpose is a key concept for privacy policies. Some policies limit the use of certain information to an explicit list of purposes. The privacy policy of The Bank of America states, "Employees are authorized to access Customer Information for business purposes only." [5]. The HIPAA Privacy Rule requires that healthcare providers in the U.S. use protected health information about a patient with that patient's authorization or only for a fixed list of allowed purposes, such as treatment and billing [30]. Other policies prohibit using certain information for a purpose. For example, Yahoo!'s privacy policy states "Yahoo!'s practice on Yahoo! Mail Classic is not to use the content of messages stored in your Yahoo! Mail account for marketing purposes." [47].

Each of these examples presents a constraint on the purposes for which the organization may use information. We call these constraints *purpose restrictions*.

Let us consider a purpose restriction in detail. As a simplification of the Yahoo! example, consider an advertising network attempting to determine which advertisement to show for marketing to a visitor of a website (such as an email website). To improve its public image and to satisfy government regulations, the network adopts a privacy policy containing a restriction prohibiting the use of the visitor's gender for the purpose of marketing.

The network has access to a database of information about potential visitors, which includes their gender. Since some advertisements are more effective, on average, for some demographics than others, using this information is in the network's interest. However, the purpose restriction prohibits the use of gender for

*This research was supported by the U.S. Army Research Office grants DAAD19-02-1-0389 and W911NF-09-1-0273 to CyLab, by the National Science Foundation (NSF) grants CCF0424422 and CNS1064688, and by the U.S. Department of Health and Human Services grant HHS 90TR0003/01. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

†The authors conducted most of this work while at Carnegie Mellon University.

selecting advertisements since it is a form of marketing. Since tension exists between selecting the most effective ad and obeying the purpose restriction, internal compliance officers and government regulators should audit the network to determine whether it has complied with the privacy policy.

However, the auditors may find manually auditing the network difficult and error prone leading them to desire automated tools to aid them. Indeed, the difficulty of manually auditing purpose restrictions has led to commercial software for this task (e.g., [14]). However, their approaches have been ad hoc.

Our goal is to place purpose restrictions governing information use on a formal footing and to automate their enforcement. In the above example, intuitively, the auditor must determine what information the network used while planning which ads to show to a user. In general, determining whether the purpose restriction was obeyed involves determining facts about how the audited agent (a person, organization, or computer system) planned its actions. In particular, philosophical inquiry [41] and an empirical study [42] show that the behavior of an audited agent is for a purpose when the agent chooses that behavior while planning to satisfy the purpose. Our prior work has used a formal model of planning to automate the auditing of purpose restrictions that limit visible actions to certain purposes [42].

We build upon that work to provide formal semantics and algorithms for purpose restrictions limiting *information uses*, whose occurrence the auditor cannot directly observe. For example, while the ad network is prohibited from using the visitor's gender, it may access the database to use other information even if the database returns the gender as part of a larger record. Thus, our model must elucidate whether the network *used* the gender component of the accessed information.

To provide auditing algorithms, we need a formal model of planning. Fortunately, research in artificial intelligence has provided a variety of formal models of planning. To select an appropriate model for auditing, we examine the key features of our motivating example of the ad network. First, it shows that purposes are not just goals to be achieved since the purpose of marketing is quantitative: marketing can be satisfied to varying degrees and more can always be done. Second, the example shows that outcomes can be probabilistic since the network does not know what ad will be best for each visitor but does have statistical information about various demographics. Lastly, the policy is governing the use of information. Thus, our model needs an explicit model of information.

The first two features suggest using Markov Decision Processes (MDPs), which we have successfully used in an auditing algorithm for purpose restrictions on observable actions [42]. However, needing an explicit model of information requires us to use an extension of MDPs, Partially Observable Markov Decision Processes (POMDPs), which make the ability of the planning agent to observe its environment and collect information explicit. We use a POMDP to model the agent's environment where the purpose in question defines the reward function of the POMDP. The explicitness of observations (inputs) in the POMDP model allows us to go beyond standard research on planning to provide a semantics of *information use* by considering how the agent would plan if some observations were conflated to ignore information of interest.

In more detail, we quotient the POMDP's space of observations to express information use. Intuitively, to use information is to see a distinction, and to not use information corresponds to ignoring this distinction. Thus, we quotient by an equivalence relation that treats two observations as indistinguishable if they differ only by information whose use is prohibited by a purpose restriction. For example, the ad network promising not to use gender should quotient its observations by an equivalence relation that treats the genders as equivalent. By conflating observations that differ only by gender, the network will ignore gender, simulating ignorance of it. Such quotienting is defined for POMDPs since observations probabilistically constrain the space of possible current states of the agent's environment, and quotienting just decreases the constraint's accuracy.

We use our quotienting operation to provide two different definitions of what it means for an agent to obey a purpose restriction involving information use. The first requires that the agent uses the quotiented POMDP to select its behavior. We call this definition *cognitive* since it refers to the agent's cognitive process of selecting behavior. Since the auditor cannot examine the agent's cognitive processes and might only care about their external consequences, we offer a second weaker definition that depends upon the agent's observable behavior. The *behaviorist* definition only requires that the agent's behaviors be consistent with using the quotiented POMDP. It does not depend upon whether the agent actually used that POMDP

or a different process to select its behavior.

We use the behaviorist definition as the basis of an auditing algorithm that compares the behaviors of an agent to each of the behaviors that is acceptable under our notion of simulated ignorance. Despite comparing to multiple behaviors, our algorithm only needs to optimize the quotiented POMDP once. For the behaviorist definition, we prove that the algorithm is sound (Theorem 1) and is complete when the POMDP can be optimized exactly (Theorem 2).

To show that our semantics is not too weak, we compare it to *noninterference*, a formalization of information use for automata found in prior security research [15]. This definition examines how an input to an automaton affects the automaton's output. Our approach is similar but uses POMDPs instead of automata. We relate the two models by defining how an automaton can implement a strategy for a quotiented POMDP, which allows us to prove that the cognitive definition implies a form of noninterference (Theorem 3). On the other hand, we show that an agent can obey the behaviorist definition while still exhibiting interference. However, interestingly, such interference cannot actually further the restricted purpose showing that the behaviorist definition is still strong enough to prevent interference *for that purpose*.

Since an action's purpose can depend upon how it fits into a chain of actions, we focus on post-hoc auditing. Nevertheless, other enforcement mechanisms can employ our semantics. Despite focusing on privacy policies, our semantics and algorithm may aid the enforcement of other policies restricting the use of information to only certain purposes, such as those governing intellectual property.

Contributions and Outline. We start by reviewing related work and POMDPs (Sections 2 and 3). Our first contribution is definitional: we use our quotienting characterization of information use to provide both the cognitive and behaviorist definitions of complying with a purpose restriction on information use (Section 4). Our second contribution is our auditing algorithm accompanied by theorems of soundness and a qualified form of completeness (Section 5). Our final contribution is relating our formalization to noninterference with a theorem showing that the cognitive definition implies noninterference (Sections 6). We end with conclusions (Sections 7).

2 Prior Work

Our work builds upon three strands of prior work: information flow analysis, enforcing purpose restrictions, and planning.

Information Flow Analysis. Research on information flow analysis led to noninterference [15], a formalization of information flow, or use. However, prior methods of detecting noninterference have typically required access to the program running the system in question. These analyses either used the program for directly analyzing its code (see [37] for a survey), for running an instrumented version of the system (e.g., [44, 28, 45, 24]), or for simulating multiple executions of the system (e.g., [48, 10, 12]). Traditionally, the requirement of access to the program has not been problematic since the analysis has been motivated as a tool for software engineers securing a program that they have designed.

However, in our setting of enforcing purpose restrictions, such access is not always possible since the analyzed system can be a person who could be adversarial and whose behavior the auditor can only observe. On the other hand, the auditor has information about the purposes that the system should be pursuing. Since the system is a purpose-driven agent, the auditor can understand its behavior in terms of a POMDP model of its environment. Thus, while prior work provides a definition of information use, it does not provide appropriate models or methods for determining whether it occurs in our setting.

Enforcing Purpose Restrictions. Most prior work on using formal methods for enforcing purpose restrictions has focused on when observable actions achieve a purpose [1, 8, 2, 9, 32, 18, 29, 13]. That is, they define an action as being for a purpose if that action (possibly as part of a chain of actions) results in that purpose being achieved. Our work differs from these works in two ways.

First, we define an action as being for a purpose when that action is part of a plan for maximizing the satisfaction of that purpose. Our definition differs by treating purposes as rewards that can be satisfied to varying degrees and by focusing on the plans rather than outcomes, which allows an action to be for a purpose even if it probabilistically fails to improve it. The semantics of purpose we use follows from informal philosophical inquiry [41] and prior work using Markov Decision Processes to formalize purpose restrictions for actions [42]. Jafari et al. offer an alternative view of planning and purposes in which a purpose is high-level action related to low-level actions by a plan [17]. Our views are complementary in that theirs picks up where ours leaves off: Our model of planning can justify the plans that their model accepts as given while their model allows for reasoning about the relationships among purposes with a logic.

Second, we consider information use. While the aforementioned works address restrictions on information *access*, they do not have a model of information *use*, such as noninterference [15]. Hayati and Abadi provide a type system for tracking information flow in programs with purpose restrictions in mind [16]. However, their work presupposes that the programmer can determine the purpose of a function and provides no formal guidance for making this determination.

Minimal disclosure requires that the amount of information used in granting a request for access should be as little as possible while still achieving the purpose behind the request. This is closely related to enforcing purpose restrictions. However, purpose restrictions do not require the amount of information used to be minimal and often involve purposes that are never fully achieved (e.g., more marketing is always possible). Unlike works on minimal disclosure [22, 6] that model purposes as conditions that are either satisfied or not, we model them as being satisfied to varying degrees. Furthermore, we model probabilistic factors absent in these works that can lead to an agent's plan failing. Modeling the such failures allows us to identify when information use is for a purpose despite not increasing the purpose's satisfaction due to issues outside of the agent's control.

Planning. Since our formal definition is in terms of planning, automating auditing depends upon automated *plan recognition* [38]. We build upon works that use models of planning to recognize plans (e.g., [4, 3, 34, 35]). The most related work has provided methods of determining when a sequence of actions are for a purpose (or "goal" in their nomenclature) given a POMDP model of the environment [35]. Our algorithm for auditing is similar to their algorithm. However, whereas their algorithm attempts to determine the probability that a sequence of actions are for a purpose, we are concerned with whether a use of information could be for a purpose. Thus, we must first develop a formalism for information use. We must also concern ourselves with the soundness of our algorithm rather than its accuracy in terms of a predicted probability. Additionally, we use traditional POMDPs to model purposes that are never fully satisfied instead of the goal POMDPs used in their work.

3 Modeling Purpose-Driven Agents

We review the Partially Observable Markov Decision Process (POMDP) model and then show how to model the above motivating example as one. We start with an agent, such as a person, organization, or artificially intelligent computer, that attempts to maximize the satisfaction of a purpose. The agent uses a POMDP to plan its actions. The POMDP models the agent's environment and how its actions affects the environment's state and the satisfaction of the purpose. The agent selects a plan that optimizes the expected total discounted reward (degree of purpose satisfaction) under the POMDP. This plan corresponds to the program running the audited system.

POMDPs. To define POMDPs, let $\mathsf{Dist}(X)$ denote the space of all distributions over the set X and let $\mathbb{R}$ be the set of real numbers. A POMDP is a tuple $\langle \mathcal{Q}, \mathcal{A}, \tau, \rho, \mathcal{O}, \nu, \gamma \rangle$ where

- $\mathcal{Q}$ is a finite state space representing the states of the agent's environment;
- $\mathcal{A}$, a finite set of actions;

- $\tau : \mathcal{Q} \times \mathcal{A} \to \mathsf{Dist}(\mathcal{Q})$, a transition function from a state and an action to a distribution over states representing the possible outcomes of the action;
- $\rho : \mathcal{Q} \times \mathcal{A} \to \mathbb{R}$, a reward function measuring the immediate impact on the satisfaction of the purpose when the agent takes the given action in the given state;
- $\mathcal{O}$, a finite observation space containing any observations the agent may perceive while performing actions;
- $\nu : \mathcal{A} \times \mathcal{Q} \to \mathsf{Dist}(\mathcal{O})$, a distribution over observations given an action and the state resulting from performing that action; and
- γ, a discount factor such that $0 \leq \gamma < 1$.

We say that a POMDP *models a purpose* if ρ measures the degree to which the purpose is satisfied. To select actions for that purpose, the agent should select those that maximizes its expected total discounted reward, $\mathbb{E}\left[\sum_{i=0}^{\infty} \gamma^i u_i\right]$ where i represents time and u_i, the reward from the agent's ith action.

This goal is complicated by the agent not knowing *a priori* which of the possible states of the POMDP is the current state of its environment. Rather it holds beliefs about which state is the current state. In particular, the agent assigns a probability to each state q according to how likely the agent believes that the current state is the state q. A *belief state* β captures these beliefs as a distribution over states of $\mathcal{Q}$ (i.e., $\beta \in \mathsf{Dist}(\mathcal{Q})$). An agent updates its belief state as it performs actions and makes observations. When an agent takes the action a and makes the observation o starting with the beliefs β, the agent develops the new beliefs β' where $\beta'(q')$ is the probability that q' is the next state.

We define $\mathsf{up}_m(\beta, a, o)$ to equal the updated beliefs β'. β' assigns to the state q' the probability $\beta'(q') = \Pr[\mathrm{Q}'{=}q' | \mathrm{O}{=}o, \mathrm{A}{=}a, \mathrm{B}{=}\beta]$ where Q' is a random variable over next states, $\mathrm{B}{=}\beta$ identifies the agent's current belief state as β, $\mathrm{A}{=}a$ identifies the agent's current action as a, and $\mathrm{O}{=}o$ identifies the observation the agent makes while performing action a as o. We may reduce $\mathsf{up}_m(\beta, a, o)$ to the following formula in terms of the POMDP model:

$$\mathsf{up}_m(\beta, a, o)(q') = \frac{\nu(a, q')(o) \sum_{q \in \mathcal{Q}} \beta(q) * \tau(q, a)(q')}{\sum_{q' \in \mathcal{Q}} \nu(a, q')(o) \sum_{q \in \mathcal{Q}} \beta(q) * \tau(q, a)(q')}$$

To maximize its expected total discounted reward, the agent does not need to track its history of actions and observations independently of its beliefs as such beliefs are a sufficient statistic. Thus, the agent need only consider for each possible belief β it can have, what action it would perform. That is, the agent can plan by selecting a *strategy*: a function from the space of beliefs $\mathsf{Dist}(\mathcal{Q})$ to the space of actions $\mathcal{A}$. (We use the word "strategy" instead of the more common "policy" to avoid confusion with privacy policies.)

The goal of the agent is find the optimal strategy. By the Bellman equation [7], the expected value of a belief state β under a strategy σ is

$$V_m(\sigma, \beta) = \mathrm{R}_m(\beta, \sigma(\beta)) + \gamma \sum_{o \in \mathcal{O}} \mathrm{N}_m(\beta, \sigma(\beta))(o) * V_m(\sigma, \mathsf{up}_m(\beta, \sigma(\beta), o)) \tag{1}$$

where R_m and N_m are ρ and ν raised to work over beliefs: $\mathrm{R}_m(\beta, a) = \sum_{q \in \mathcal{Q}} \beta(q) * \rho(q, a)$ and $\mathrm{N}_m(\beta, a)(o) = \sum_{q, q' \in \mathcal{Q}} \beta(q) * \tau(q, a)(q') * \nu(a, q')(o)$. A strategy σ is optimal if it maximizes V_m for all belief states, that is, if for all β, $V_m(\sigma, \beta)$ is equal to $V_m^*(\beta) = \max_{\sigma'} V_m(\sigma', \beta)$. Prior work has provided algorithms for finding optimal strategies by reducing the problem to one of finding an optimal strategy for a related Markov Decision Process (MDP) that uses these belief states as its state space (e.g., [40]). (For a survey, see [27].)

Example. We can formalize the motivating example provided in Section 1 as a POMDP m_{ex}. Here, we provide an overview that is sufficient for understanding the rest of the paper; the appendix provides additional details.

For simplicity, we assume that the only information relevant to advertising is the gender of the visitor. Thus, the state space $\mathcal{Q}$ is determined by three factors: the visitor's gender, the gender (if any) recorded in the database, and what advertisement (if any) the network has shown to the visitor.

Also for simplicity, we assume that the network is choosing among three advertisements. We use the action space $\mathcal{A} = \{\mathsf{lookup}, \mathsf{ad}_1, \mathsf{ad}_2, \mathsf{ad}_3\}$. The actions ad_1, ad_2, and ad_3 correspond to the network showing the visitor one of the three possible advertisements while lookup corresponds to the network looking up information on the visitor. We presume ad_1 is the best for females and the worst for males, ad_3 is the best for males and the worst for females, and ad_2 strikes a middle ground. In particular, we use $\rho(q, \mathsf{ad}_1) = 9$ for a state q in which the visitor is a female and has not yet seen an ad. The reward 9 could refer to a measure of the click through rate or the average preference assigned to the ad by females during market research. If the visitor were instead a male, the reward would be 3. For ad_3, the rewards are reversed with 3 for females and 9 for males. For ad_2, the reward is 7 for both genders. The action lookup or showing a second ad produces reward of zero. We use a discounting factor of $\gamma = 0.9$.

The function τ shows how actions change the environment's state while ν shows how observations accompany these actions. τ enforces that showing an ad changes the state into one in which showing a second ad produces no further rewards. It also specifies that performing lookup does not change the state of the environment. On the other hand, ν shows that lookup can change the state of the agent's knowledge. In particular, it shows that performing lookup produces an observation $\langle d, \alpha \rangle$. The observation reveals that the database holds data d about the visitor's gender and α about what if any ad the visitor has seen. Thus, the observation space is $\mathcal{O} = \{\mathsf{f}, \mathsf{m}, \bot\} \times \{\mathsf{ad}_1, \mathsf{ad}_2, \mathsf{ad}_3, \varnothing\}$ with f for the database showing a female, m for a male, $\bot$ for no gender entry, ad_i for the visitor having seen ad_i, and $\varnothing$ for the visitor having not seen an ad.

How the network will behave depends upon the network's initial beliefs β_{ex1}. We presume that the network believes its database's entries to be correct, that it has not shown an advertisement to the visitor yet, and that visitors are equally likely to be female or male. Under these assumptions, the optimal plan for the network is to first check whether the database contains information about the visitor. If the database records that the visitor is a female, then the network shows her ad_1. If it records a male, the network shows ad_3. If the database does not contain the visitor's gender (holds $\bot$), then the network shows ad_2. The optimal plan is not constrained as to what the agent does after showing the advertisement as it does not affect the reward. (We return to this point later when we consider non-redundancy in Section 5.)

This optimal plan characterizes the form of the set of optimal strategies. The set contains multiple optimal strategies since the network is unconstrained in the actions it performs after showing the advertisement. The optimal strategies must also specify how the network would behave under other possible beliefs it could have had. For example, if the network believed that all visitors are females regardless of what its database records, then it would always show ad_1 without first checking its database.

Intuitively, using any of these optimal strategies would violate the privacy policy prohibiting using gender for marketing. The reason is that the network selected which advertisement to show using the database's information about the visitor's gender.

We expect the network constrained to obeying the policy will show ad_2 to all visitors (presuming approximately equal numbers of female and male visitors). Our reasoning is that the network must plan as though it does not know and cannot learn the visitor's gender. In this state of simulated ignorance, the best plan the network can select is the middle ground of ad_2. The next section formalizes this planning under simulated ignorance.

4 Constraining POMDPs for Information Use

We now provide a formal characterization of how an agent pursuing a purpose should behave when prohibited from using a class of information. Recall the intuition that using information is using a distinction and that not using it corresponds to ignoring the distinction. We use this idea to model sensitive information with an equivalence relation $\equiv$. We set $o_1 \equiv o_2$ for any two observations o_1 and o_2 that differ only by sensitive information.

From $\equiv$ and a POMDP m, we construct a POMDP $m/\equiv$ that ignores the prohibited information. For each equivalence class of $\equiv$, $m/\equiv$ will conflate its members by treating every observation in it as indistinguishable from one another. To ignore these distinctions, on observing o, the agent updates its belief state as though it has seen some element of $\equiv[o]$ but is unsure of which one where $\equiv[o]$ is the equivalence class that holds

the observation o.

To make this formal, we define a quotient POMDP $m/\equiv$ that uses a quotiented space of observations. Let $\mathcal{O}/\equiv$ be the set of equivalence classes of $\mathcal{O}$ under $\equiv$. Let $\nu/\equiv$ give the probability of seeing any observation of an equivalence class: $\nu/\equiv(a,q')(O) = \sum_{o\in O}\nu(a,q')(o)$ where O is an equivalence class in $\mathcal{O}/\equiv$. Given $m = \langle \mathcal{Q}, \mathcal{A}, \tau, \rho, \mathcal{O}, \nu, \gamma\rangle$, let $m/\equiv$ be $\langle \mathcal{Q}, \mathcal{A}, \tau, \rho, \mathcal{O}/\equiv, \nu/\equiv, \gamma\rangle$.

Proposition 1. *For all POMDPs m and equivalences $\equiv$, $m/\equiv$ is a POMDP.*

Proof. We prove that $\nu/\equiv$ produces probability distributions over $\mathcal{O}/\equiv$. For all a and q',

$$\sum_{O\in\mathcal{O}/\equiv}\nu/\equiv(a,q')(O) = \sum_{O\in\mathcal{O}/\equiv}\sum_{o\in O}\nu(a,q')(o) = \sum_{o\in\mathcal{O}}\nu(a,q')(o) = 1$$

follows from $\mathcal{O}/\equiv$ being a partition of $\mathcal{O}$ and from $\nu(a,q')$ being a distribution over $\mathcal{O}$. For all $O\in\mathcal{O}/\equiv$, $0 \le \nu/\equiv(a,q')(O) \le 1$ since $\nu/\equiv(a,q')(O) = \sum_{o\in O}\nu(a,q')(o)$ and $O\subseteq\mathcal{O}$. Thus, $\nu/\equiv(a,q')$ is a probability distribution. □

Example. Returning to the example POMDP of Section 3, the policy governing the network states that the network will not use the database's entry about the visitor's gender for determining the advertisement to show the visitor. The auditor must decide how to formally model this restriction. One way would be to define $\equiv_{\mathsf{ex}}$ such that for all g and g' in $\{\mathsf{f},\mathsf{m},\bot\}$, and α in $\{\mathsf{ad}_1,\mathsf{ad}_2,\mathsf{ad}_3,\varnothing\}$, $\langle g,\alpha\rangle \equiv_{\mathsf{ex}} \langle g',\alpha\rangle$, conflating the gender for all observations. Under this requirement, $m_{\mathsf{ex}}/\equiv_{\mathsf{ex}}$ will be such that the optimal strategy will be determined solely by the network's initial beliefs and performing the action lookup will be of no benefit. Any optimal strategy for $m_{\mathsf{ex}}/\equiv_{\mathsf{ex}}$ will call for performing ad_2 from the initial beliefs β_{ex1} discussed above.

Alternatively, the auditor might conclude that the policy only forces the network to ignore whether the database records the visitor as a female or male and not whether the database contains this information. In this case, the auditor would use a different equivalence $\equiv'_{\mathsf{ex}}$ such that $\langle\mathsf{f},\alpha\rangle \equiv'_{\mathsf{ex}} \langle\mathsf{m},\alpha\rangle$ but $\langle\mathsf{f},\alpha\rangle \not\equiv'_{\mathsf{ex}} \langle\bot,\alpha\rangle \not\equiv'_{\mathsf{ex}} \langle\mathsf{m},\alpha\rangle$ for all α. Under the initial beliefs β_{ex1}, the network would behave identically under $\equiv_{\mathsf{ex}}$ and $\equiv'_{\mathsf{ex}}$. However, if the network's beliefs were such that it is much more likely to not know a female's gender than a male's, then it might choose to show ad_1 instead of ad_2 in the case of observing $\langle\bot,\varnothing\rangle$.

The next proposition proves that we constructed the POMDP $m/\equiv$ so that beliefs are updated as if the agent only learns that some element of an equivalence class of observations was observed but not which one. That is, we prove that the updated belief $\mathsf{up}_{m/\equiv}(\beta,a,\equiv[o])(q')$ is equal to the probability that the next environmental state is q' given the distribution β over possible last states, that the last action was a, and that the observation was a member of $\equiv[o]$. Recall that Q′ is a random variable over the next state while O, A, and B identify the last observation, action, and belief state, respectively.

Proposition 2. *For all POMDPs m, equivalences $\equiv$, beliefs β, actions a, observations o, and states q', $\mathsf{up}_{m/\equiv}(\beta,a,\equiv[o])(q') = \Pr[\mathrm{Q}'{=}q' \mid \mathrm{O}\in\equiv[o], \mathrm{A}{=}a, \mathrm{B}{=}\beta]$.*

Proof. For all m, $\equiv$, β, a, o, and q',

$$\begin{aligned}
\mathsf{up}_{m/\equiv}(\beta,a,\equiv[o])(q') &= \frac{\nu/\equiv(a,q')(\equiv[o])\sum_{q\in\mathcal{Q}}\beta(q)*\tau(q,a)(q')}{\sum_{q'\in Q}\nu/\equiv(a,q')(\equiv[o])\sum_{q\in Q}\beta(q)*\tau(q,a)(q')}\\
&= \frac{\sum_{o_1\in\equiv[o]}\nu(a,q')(o_1)\sum_{q\in\mathcal{Q}}\beta(q)*\tau(q,a)(q')}{\sum_{o_1\in\equiv[o]}\sum_{q'\in Q}\nu(a,q')(o_1)\sum_{q\in Q}\beta(q)*\tau(q,a)(q')}\\
&= \frac{\Pr[O\in\equiv[o]|Q'{=}q', A{=}a, B{=}\beta]\Pr[Q'{=}q'|A{=}a,B{=}\beta]}{\Pr[O\in\equiv[o]|A{=}a,B{=}\beta]}\\
&= \Pr[\mathrm{Q}'{=}q' \mid \mathrm{O}\in\equiv[o], \mathrm{A}{=}a, \mathrm{B}{=}\beta]
\end{aligned}$$

since $\Pr[O\in\equiv[o]|Q'{=}q', A{=}a, B{=}\beta] = \Pr[O\in\equiv[o]|Q'{=}q', A{=}a]$. □

Propositions 1 and 2 show that $m/\equiv$ is a POMDP that ignores the distinctions among observations that only differ by sensitive information. They justify the following definition, which explains how a purpose-driven agent should act when prohibited from using certain information. They show that it correctly prevents the use of the prohibited information. The definition's appeal to optimizing a POMDP is justified by our prior work showing that an action is for a purpose when that action is selected as part of a plan optimizing the satisfaction of that purpose [42]. We extend this result to information by concluding that information used to select an action is used for that action's purpose.

Definition 1 (Cognitive)**.** *An agent* obeys *the purpose restriction to perform actions for the purpose modeled by the POMDP m without using the information modeled by $\equiv$ iff the agent selects an strategy by optimizing $m/\equiv$.*

We call the above definition *cognitive* since it refers to the strategy selected by the agent as part of a cognitive process that the auditor cannot measure. Rather, the auditor can only view the agent's external behavior and visible aspects of the environment. That is, the auditor can only view the agent's actions and observations, which we refer to collectively as the agent's *execution*.

We can formalize the agent's execution using a function exe. Even when the agent uses the POMDP $m/\equiv$ with observation space $\mathcal{O}/\equiv$ to select a strategy, the actual observations the agent makes lie in $\mathcal{O}$, complicating exe. We recursively define $\mathsf{exe}(m, \equiv, \sigma, \beta_1, \vec{o})$ to be the agent's execution that arises from it employing a strategy σ observing a sequence of observations $\vec{o} = [o_1, \ldots o_n]$ in $\mathcal{O}^*$ starting with beliefs β_1 for a POMDP $m/\equiv$. For the empty sequence [] of observations, $\mathsf{exe}(m, \equiv, \sigma, \beta, []) = [\sigma(\beta)]$ since the agent can only make one action before needing to wait for the next observation and updating its beliefs. For non-empty sequences $o{:}\vec{o}$, it is equal to $\sigma(\beta){:}o{:}\mathsf{exe}(m, \equiv, \sigma, \mathsf{up}_{m/\equiv}(\beta, \sigma(\beta), \equiv[o]), \vec{o})$ where $x{:}y$ denotes prepending element x to the sequence y.

A single execution $\vec{e}$ can be consistent with both an optimal strategy for $m/\equiv$ and a strategy that is not optimal for $m/\equiv$. Consider for example, the execution $\vec{e} = [\mathsf{ad}_2] = \mathsf{exe}(m_{\mathsf{ex}}, \equiv_{\mathsf{ex}}, \sigma, \beta_{\mathsf{ex}}, [])$ that arises from an optimal strategy σ for $m_{\mathsf{ex}}/\equiv_{\mathsf{ex}}$. This execution can also arise from the agent planning for a different purpose, such as maximizing kickbacks for showing certain ads, provided that ad_2 also just so happens to maximize that purpose. Since the auditor only observes the execution $\vec{e}$ and not the cognitive process that selected the action ad_2, the auditor cannot know by which process the agent selected the ad. Thus, the auditor cannot determine from an execution that an agent obeyed a purpose restriction under Definition 1.

Some auditors may find this fundamental limitation immaterial since such an agent's actions are still consistent with an allowed strategy. Since the actual reasons behind the agent selecting those actions do not affect the environment, an auditor might not find concerning an agent doing the right actions for the wrong reasons. To capture this more consequentialist view of compliance, we provide a weaker definition that focuses on only the agent's execution.

Definition 2 (Behaviorist)**.** *An agent performing execution $\vec{e}$* obeys *the purpose restriction to perform actions for the purpose modeled by the POMDP m and initial beliefs β_1 without using the information modeled by the equivalence relation $\equiv$ given the observations $\vec{o}$ iff $\vec{e} = \mathsf{exe}(m, \equiv, \sigma, \beta_1, \vec{o})$ for some σ that is an optimal strategy of $m/\equiv$.*

5 Auditing Algorithm

Under the behaviorist definition, to determine whether an agent obeyed a prohibition against using certain information for a purpose pursued by the agent, the auditor can compare the agent's behaviors to the appropriate strategies. The auditor records the agent's execution in a log ℓ that shows the actions and observations of the agent. For example, databases for electronic medical records log many of the actions and observations of healthcare providers. The auditor may then compare the recorded behavior to that dictated by Definition 2, i.e., to the optimal strategies for the quotient POMDP modeling the purpose while ignoring disallowed information.

Given our formal model, we can automate the comparison of the agent's behavior to the allowable behavior. We use an algorithm Audit that takes as inputs a POMDP m, an equivalence relation $\equiv$, and

a log $\ell = [a_1, o_1, a_2, o_2, \ldots, a_n, o_n]$ such that the audited agent is operating in the environment m under a policy prohibiting information as described by $\equiv$ and took action a_i followed by observation o_i for all $i \leq n$. For simplicity, we assume that ℓ records all relevant actions and observations. AUDIT returns whether the agent's behavior, as recorded in ℓ, is inconsistent with optimizing the POMDP $m/\equiv$.

AUDIT operates by first constructing the quotient POMDP $m/\equiv$ from m and $\equiv$. Next, similar to a prior algorithm [35], for each i, AUDIT checks whether performing the recorded action a_i in the current belief state β_i is optimal under $m/\equiv$. The algorithm constructs these belief states from the observations and initial belief state β_1. Due to the complexity of solving POMDPs [31], we use an approximation algorithm to solve for the value of performing a_i in β_i (denoted $Q^*_{m/\equiv}(\beta_i, a_i)$) and the optimal value $V^*_{m/\equiv}(\beta_i)$. Unlike prior work, for soundness, we require an approximation algorithm SOLVEPOMDP that produces both lower bounds $\mathtt{V}^*_{\mathsf{low}}$ and upper bounds $\mathtt{V}^*_{\mathsf{up}}$ on $V^*_{m/\equiv}(\beta_i)$. Many such algorithms exist (e.g., [49, 39, 20, 33]). For each β_i and a_i in ℓ, AUDIT checks whether these bounds show that $Q^*_{m/\equiv}(\beta_i, a_i)$ is strictly less than $V^*_{m/\equiv}(\beta_i)$. If so, then the action a_i is sub-optimal for β_i and AUDIT returns true. Pseudo-code for AUDIT follows:

AUDIT($\langle \mathcal{Q}, \mathcal{A}, \tau, \rho, \mathcal{O}, \nu, \gamma \rangle, \equiv, \beta_1, [a_1, o_1, a_2, o_2, \ldots, a_n, o_n]$):
01 $m' = \langle \mathcal{Q}, \mathcal{A}, \tau, \rho, \mathcal{O}/\equiv, \nu/\equiv, \gamma \rangle$
02 $\langle \mathtt{V}^*_{\mathsf{low}}, \mathtt{V}^*_{\mathsf{up}} \rangle :=$ SOLVEPOMDP(m')
03 for ($i := 1$; $i \leq n$; i++):
04 if ($\mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta_i, a_i) < \mathtt{V}^*_{\mathsf{low}}(\beta_i)$):
05 return true
06 $\beta_{i+1} := \mathsf{up}_{m/\equiv}(\beta_i, a_i, \equiv[o_i])$;
07 return false

where $\mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta, a)$ is a function that uses $\mathtt{V}^*_{\mathsf{up}}$ to return an upper bound on $Q^*_{m/\equiv}(\beta, a)$:

$$\mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta, a) = \mathrm{R}_m(\beta, a) + \gamma \sum_{O \in \mathcal{O}/\equiv} \mathrm{N}_m(\beta, a))(O) * \mathtt{V}^*_{\mathsf{up}}(\mathsf{up}_{m'}(\beta, \sigma(\beta), O))$$

Theorem 1 (Soundness)**.** *If* AUDIT *returns true, then the agent did not follow an optimal strategy for $m/\equiv$, violating both Definitions 1 and 2.*

Proof. If the algorithm returns true, then for some i, $Q^*_{m/\equiv}(\beta_i, a_i) \leq \mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta_i, a_i) < \mathtt{V}^*_{\mathsf{low}}(\beta_i) \leq V^*_{m/\equiv}(\beta_i)$. This implies that a_i is suboptimal at belief state β_i and the agent did not follow an optimal strategy for the allowed purpose using only the allowed information. □

Thus, if AUDIT returns true, either the agent optimized some other purpose, used information it should not have, used a different POMDP model of its environment, or failed to correctly optimize the POMDP. Each of these possibilities should concern the auditor and is worthy of further investigation.

If the algorithm returns false, then the auditor cannot find the agent's behavior inconsistent with an optimal strategy and should spend his time auditing other agents. However, AUDIT is incomplete and such a finding does not mean that the agent surely performed its actions for the purpose without using the prohibited information. For the cognitive definition, incompleteness is unavoidable since the definition depends upon cognitive constructs that the auditor cannot measure. For example, recall that the network could display the execution $\vec{e} = [\mathsf{ad}_2]$ either from performing the allowed optimization or by performing some disallowed optimization that also results in the action ad_2 being optimal.

For the behaviorist definition, incompleteness results since a better approximation might actually show that $Q^*_{m/\equiv}(\beta_i, a_i) < V^*_{m/\equiv}(\beta_i)$ for some i. In principle this source is avoidable by using an exact POMDP solver instead of an approximate one. However, the exact solution to some POMDPs is undecidable [21]. Nevertheless, we can prove that this inability is the only source of incompleteness.

Theorem 2 (Qualified Completeness)**.** *If* Audit *using an oracle to exactly solve POMDPs returns false, then the agent obeyed the purpose restriction according to the behaviorist definition (Definition 2).*

Proof. Assume that algorithm returns false. Then, for every i, it must be the case that $\mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta_i, a_i) \not< \mathtt{V}^*_{\mathsf{low}}(\beta_i)$. Since an oracle returns exact results for $\mathtt{V}^*_{\mathsf{up}}$ and $\mathtt{V}^*_{\mathsf{low}}$, $Q^*_{m/\equiv}(\beta_i, a_i) = \mathtt{Q}^*_{\mathsf{up}}(\mathtt{V}^*_{\mathsf{up}}, \beta_i, a_i)$ and $\mathtt{V}^*_{\mathsf{low}}(\beta_i) =$

$V^*_{m/\equiv}(\beta_i)$. Thus, for all i, $Q^*_{m/\equiv}(\beta_i, a_i) \geq V^*_{m/\equiv}(\beta_i)$. Thus for all i, a_i is optimal at belief state β_i and the agent's are consistent with following an optimal strategy for $m/\equiv$. □

Other Purpose Restrictions. AUDIT is specialized for determining whether or not the audited agent performed its actions for a purpose without using some prohibited information. While such a question is relevant to an internal compliance officer auditing employees, it does not correspond to the purpose restrictions found in outward-facing privacy policies.

One type of restriction found in such policies is the *not-for* restriction prohibiting information from being used for a purpose. For example, Yahoo! promised to *not* use contents of emails *for* marketing. This restriction is similar to the condition checked by AUDIT, but is weaker in that audited agent may obey it either (1) by performing actions for that purpose without using that information (which AUDIT checks) or (2) by not performing actions for that purpose.

A second type is the *only-for* restriction, which limits the agent to using a class of information only for a purpose. For example, HIPAA requires that medical records are used *only for* certain purposes such as treatment. It is also weak in that the agent can obey it either (1) by performing actions for the purpose (which AUDIT checks using equality for $\equiv$ to allow the agent to use the information) or (2) by not using the information in question while performing actions for some other purpose.

For both of these types, our algorithm can handle the first option (1) for compliance. However, for both these types, the second option (2) for compliance involves an open-ended space of possible alternative purposes that could have motivated the agent's actions. In some cases (e.g., healthcare), this space may be small enough to check each alternative (e.g., treatment, billing, research, training) with AUDIT. In other cases, the auditor might have the authority to compel the agent to explain what its purpose was. In either of these cases, the auditor could use AUDIT to explore these alternative purposes.

Modeling. AUDIT requires a POMDP that models how various actions affect the purpose in question. In some cases, acquiring such a model may be non-trivial. We hope that future work can ease the process of model construction using techniques from reinforcement learning, such as SARSA [36], that automatically construct models from observing the behavior of multiple agents.

In some cases, the auditor might be able to compel the agent to provide the POMDP used. In this case, AUDIT would check whether the agent's story is consistent with its actions.

Non-Redundancy. In our running example, the actions of the agent after showing the advertisement are unconstrained. The reason is that showing the advertisement will result in the current state of the POMDP becoming one from which no further rewards are possible. Since the only criterion of an optimal strategy is its expected total discounted reward, a strategy may assign any action to these states without changing whether it is optimal. However, none of the actions in $\mathcal{A}$ actually improves the satisfaction of the purpose. Thus, intuitively, the agent should just stop instead of performing any of them.

Prior work has formalized this intuition for MDPs using the idea of *non-redundancy* [42]. We may apply the same idea to POMDPs. We add to each POMDP a distinguished action stop that indicates that the agent stops and does nothing more (for the purpose in question). The stop action always produces zero reward and results in no state change: $\rho(q, \mathsf{stop}) = 0$ and $\tau(q, \mathsf{stop}) = \delta(q)$ for all q in $\mathcal{Q}$. An action a other than stop from a belief state β is *redundant* if it is no better than stopping: $Q^*_{m/\equiv}(\beta, a) \leq Q^*_{m/\equiv}(\beta, \mathsf{stop}) = 0$. A strategy is *non-redundant* if it never requires a redundant action from any belief state. We require that the strategy that the agent selects is not just optimal for the total expected discounted reward, but also that it is non-redundant.

We modify AUDIT to enforce this requirement by additionally checking that $\mathtt{Q}^*_{\mathsf{up}}(\beta_i, a_i) > 0$ for each pair of a belief state β_i and an action a_i other than stop in the log ℓ. If not, AUDIT has found a redundant action a_i indicating a violation and returns true.

6 Relationship with Noninterference

We have provided a definition of information use in terms of a POMDP. Prior work provides the *noninterference* definition of information use for automata [15]. In this section, we show that our definition implies a form of noninterference. In particular, we show that agents using strategies optimizing $m/\equiv$ has noninterference for $\equiv$, which suggests that our definition is sufficiently strong to rule out information use. We start by reviewing automata and noninterference.

Automaton Model of Systems. The agent using the POMDP to select a strategy can implement that strategy as a *control system* or *controller* (e.g., [19]). We follow Goguen and Meseguer's work and model systems as deterministic automata [15]. However, since we do not analyze the internal structure of systems (it is unavailable to the auditor), our approach can be applied to other models. We limit our discussion to deterministic systems since there are many competing generalizations of noninterference to the nondeterministic setting (e.g., [25, 46, 26]), but the main competitors collapse into standard noninterference in the deterministic case [11].

A system automaton $s = \langle t, r\rangle$ consists of a labeled transition system (LTS) t and a current state r. An LTS $t = \langle \mathcal{R}, \mathcal{O}, \mathcal{A}, \mathsf{next}, \mathsf{act}\rangle$ describes the automaton's behavior where $\mathcal{R}$ is a set of states; $\mathcal{O}$, a set of observations (inputs); $\mathcal{A}$, a set of actions (outputs); $\mathsf{next} : \mathcal{R} \times \mathcal{O} \to \mathcal{R}$ is a transition function; and $\mathsf{act} : \mathcal{R} \to \mathcal{A}$ is a function identifying the action that the automation selects given its current state. The current state $r \in \mathcal{R}$ changes as the system makes observations and takes actions.

As with POMDPs, an execution of a system s modeled as an automaton corresponds to an interleaving of observations from the environment and actions taken by the system. Let $\mathsf{exe}(s, \vec{o})$ denote the execution of s on a sequence $\vec{o}$ of observations. As for POMDPs, we define exe for systems recursively: $\mathsf{exe}(\langle t, r\rangle, []) = [\mathsf{act}(r)]$ and $\mathsf{exe}(\langle t, r\rangle, o{:}\vec{o}) = \mathsf{act}(r){:}o{:}\mathsf{exe}(\langle r, \mathsf{next}(r, o)\rangle, \vec{o})$ where $t = \langle \mathcal{R}, \mathcal{O}, \mathcal{A}, \mathsf{next}, \mathsf{act}\rangle$.

Noninterference. Recall that we set $o_1 \equiv o_2$ for any two observations o_1 and o_2 that differ only by sensitive information. To not use the sensitive information, the system s should treat such related observations identically.

To formalize this notion, we raise $\equiv$ to work over sequences of observations and actions (i.e., executions and sequences of observations). For such sequences $\vec{x}$ and $\vec{y}$ in $(\mathcal{O} \cup \mathcal{A})^*$, $\vec{x} \equiv \vec{y}$ iff they are of the same length and for each pair of elements x and y at the same position in $\vec{x}$ and $\vec{y}$, respectively, $x \equiv y$ where $\equiv$ is treated as equality when comparing actions.

Definition 3. *A system s has* noninterference *for $\equiv$ iff for all observation sequences $\vec{o}_1$ and $\vec{o}_2$ in $\mathcal{O}^*$, $\vec{o}_1 \equiv \vec{o}_2$ implies that $\mathsf{exe}(s, \vec{o}_1) \equiv \mathsf{exe}(s, \vec{o}_2)$.*

Our definition corresponds to the form of noninterference enforced by most type systems for information flow. (See [37] for a survey.) Unlike Goguen and Meseguer's definition, ours does not require the system's behavior to remain unchanged regardless of whether or not it receives sensitive information. Rather, the system's behavior may change upon receiving sensitive information, but this change must be the same regardless of the value of the sensitive information. (See [43] for a discussion.)

Relationship. We now characterize the relationship between our quotienting definition of information use and noninterference. We do so by considering a control system s operating in an environment modeled by a POMDP m. We require that s and m share the same sets of actions $\mathcal{A}$ and observations $\mathcal{O}$. However, the state spaces $\mathcal{R}$ of s and $\mathcal{Q}$ of m differ with $\mathcal{R}$ representing the internal states of the system and $\mathcal{Q}$ representing the external states of the environment.

We relate systems and strategies by saying that a system s *implements* a strategy σ for $m/\equiv$ and beliefs β_1 iff for all $\vec{o}$ in $\mathcal{O}^*$, $\mathsf{exe}(s, \vec{o}) = \mathsf{exe}(m, \equiv, \sigma, \beta_1, \vec{o})$. We denote the set of such implementing systems as $\mathsf{Imp}(m, \equiv, \sigma, \beta_1)$. This definition allows us to formalize the intuition that agents using strategies optimizing $m/\equiv$ has noninterference for $\equiv$. In fact, systems implementing any strategy for $m/\equiv$ has noninterference since any such implementation respects $\equiv$.

Theorem 3. *For all systems q, POMDPs m, initial beliefs β_1, strategies σ, and equivalences $\equiv$, if s is in $\mathsf{Imp}(m, \equiv, \sigma, \beta_1)$, then s has noninterference for $\equiv$.*

Proof. Assume that the system s is in $\mathsf{Imp}(m, \equiv, \sigma, \beta)$. Then for any observation $\vec{o}$, $\mathsf{exe}(s, \vec{o}) = \mathsf{exe}(m, \equiv, \sigma, \beta, \vec{o})$.

Suppose that $\vec{o}_1 \equiv \vec{o}_2$. Since $\vec{o}_1 \equiv \vec{o}_2$, $|\vec{o}_1| = |\vec{o}_2|$. We can prove by induction over this length that $\mathsf{exe}(m, \equiv, \sigma, \beta, \vec{o}_1) \equiv \mathsf{exe}(m, \equiv, \sigma, \beta, \vec{o}_2)$:

- Base Case: $\vec{o}_1 = []$ and $\vec{o}_2 = []$. The result follows immediately since $\vec{o}_1 = \vec{o}_2$.
- Inductive Case: $\vec{o}_1 = o_1{:}\vec{o}'_1$ and $\vec{o}_2 = o_2{:}\vec{o}'_2$. Since $\vec{o}_1 \equiv \vec{o}_2$, $\vec{o}'_1 \equiv \vec{o}'_2$ and $o_1 \equiv o_2$. For some β', $\mathsf{up}_{m/\equiv}(\beta, \sigma(\beta), \equiv[o_1]) = \beta' = \mathsf{up}_{m/\equiv}(\beta, \sigma(\beta), \equiv[o_2])$ since $o_1 \equiv o_2$. By the inductive hypothesis on $\vec{o}'_1$ and $\vec{o}'_2$, $\mathsf{exe}(m, \equiv, \sigma, \beta', \vec{o}'_1) \equiv \mathsf{exe}(m, \equiv, \sigma, \beta', \vec{o}'_2)$. Thus,

  $$\mathsf{exe}(m, \equiv, \sigma, \beta, o_1{:}\vec{o}_1) = \sigma(\beta){:}o_1{:}\mathsf{exe}(m, \equiv, \sigma, \beta', \vec{o}'_1) \equiv \sigma(\beta){:}o_2{:}\mathsf{exe}(m, \equiv, \sigma, \beta', \vec{o}'_2) = \mathsf{exe}(m, \equiv, \sigma, \beta, o_2{:}\vec{o}_2)$$

Since $\mathsf{exe}(m, \equiv, \sigma, \beta, \vec{o}_1) \equiv \mathsf{exe}(m, \equiv, \sigma, \beta, \vec{o}_2)$, $\mathsf{exe}(s, \vec{o}_1) \equiv \mathsf{exe}(s, \vec{o}_2)$. □

Agents obeying a purpose restriction under the cognitive definition (Definition 1) will employ a system in $\mathsf{Imp}(m, \equiv, \sigma, \beta_1)$. Thus, Theorem 3 shows that the cognitive definition is sufficiently strong to rule out information use.

Information Use for Other Purposes. The situation is subtler for the weaker behaviorist definition (Definition 2) and the algorithm Audit based upon it. Systems exist that will pass Audit and satisfy the behaviorist definition despite having interference by using the protected information for some purpose other than the restricted one. The key is that there could be more than one optimal strategy for a POMDP and that the agent may use the choice among optimal strategies to communicate information. The behavior of such a system will be consistent with whichever optimal strategy it selects, satisfying the behaviorist definition and Audit. However, such a system will not actually implement any strategy for the quotiented POMDP $m/\equiv$ since it distinguishes between observations conflated by $\equiv$.

For example, consider modifying the motivating example found in Section 3 in two ways to make the POMDP m'_{ex}. First, let ad_2 come in two versions, ad_2^a and ad_2^b, which are otherwise the same as the original ad_2. Second, change the POMDP so that the network must perform the action lookup before showing any ads. Two optimal non-redundant strategies will exist for $m'_{\mathsf{ex}}/\equiv$. Starting from the initial beliefs β_{ex1} discussed above, in one of the strategies, σ^a, the network will first perform lookup and then show ad_2^a. Under the second, σ^b, it will show ad_2^b after lookup. Under both, it then switches to the action stop.

The network's ability to choose between σ^a and σ^b can result in interference. In particular, the network might not implement either of them and instead delay the choice between ad_2^a and ad_2^b until after the observation from lookup informs it of the visitor's gender. The network could then use ad_2^a for a female and ad_2^b for a male. While such a system would use the information and have interference, it obeys the behaviorist definition with its actions consistent with either σ^a in the case of a female or σ^b in the case of a male.

Since such systems use the prohibited information to choose between *optimal* strategies, doing so does not actually increase its satisfaction of the purpose. Thus, this information use is not intuitively for that purpose. The agent must be motivated by some other purpose such as exfiltrating protected information to a third-party that can see which ad the network selects but the not visitor's gender directly. Thus, the behaviorist definition does not allow the agent to use the information for the purpose prohibited by the restriction, but rather allows the agent to use the information for some other purpose.

The auditor might want to prevent such interference since it violates the cognitive definition. The modifications to the example illustrate two ways that the auditor can do so if he has sufficient control over the agent's environment. The first is to ensure that only a single strategy is optimal and non-redundant. The second is to make sure that the agent can avoid learning the protected information (such as by performing the action lookup) and that learning it incurs a cost. When learning information is optional and costly, the agent will only be able to learn it if doing so increases its total reward, and not just to select among optimal

strategies that do not depend upon using that information. A third possible modification is to require the agent to perform an action committing it to a single strategy before it can learn the protected information.

In some cases an auditor can detect such information flows without modifying the POMDP. For example, intuitively, we would expect the ad network to handle more than one visitor. The auditor could compare the network's behavior when given a female to that when given a male. A difference in treatment indicates that the network is not consistently implementing either of the optimal strategies.

7 Conclusion

We use planning to create the first formal semantics for determining when information is used for a purpose. We have provided an auditing algorithm based on our formalism. We have discussed applying our algorithm to the problem of enforcing purpose restrictions found in privacy policies.

Our methods have applications beyond enforcing purpose restrictions. For example, due to privacy concerns, much interest exists in determining how third-party data collection agencies use the information they collect. (See [23] for a survey.) Despite being a question of information flow, program analyses are inapplicable since the programs are unavailable, as in our setting. Unlike our setting, these agencies typically do not subject themselves to purpose restrictions. Nevertheless, their desire for profit implicitly restrains their behavior in a manner similar to a purpose restriction. Thus, our semantics and algorithm provide a starting point for investigating such agencies.

Acknowledgments. We appreciate the discussions we have had with Lorrie Faith Cranor, Joseph Y. Halpern, and Manuela M. Veloso on this work. We thank Amit Datta, Dilsun Kaynar, and Divya Sharma for many helpful comments on this paper.

References

[1] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. Hippocratic databases. In *VLDB '02: Proceedings of the 28th International Conference on Very Large Data Bases* (2002), VLDB Endowment, pp. 143–154.

[2] Al-Fedaghi, S. S. Beyond purpose-based privacy access control. In *Proceedings of the Eighteenth Australasian Database Conference* (2007), Australian Computer Society, Inc., pp. 23–32.

[3] Baker, C. L., Saxe, R., and Tenenbaum, J. B. Action understanding as inverse planning. *Cognition 113*, 3 (2009), 329–349.

[4] Baker, C. L., Tenenbaum, J. B., and Saxe, R. R. Bayesian models of human action understanding. In *Advances in Neural Information Processing Systems 18* (2006), MIT Press, pp. 99–106.

[5] Bank of America Corp. Bank of America privacy policy for consumers, 2005.

[6] Barth, A., Mitchell, J., Datta, A., and Sundaram, S. Privacy and utility in business processes. In *CSF '07: Proceedings of the 20th IEEE Computer Security Foundations Symposium* (2007), pp. 279–294.

[7] Bellman, R. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences 38* (1952), 716–719.

[8] Byun, J.-W., Bertino, E., and Li, N. Purpose based access control of complex data for privacy protection. In *SACMAT '05: Proceedings of the Tenth ACM Symposium on Access Control Models and Technologies* (2005), ACM, pp. 102–110.

[9] Byun, J.-W., and Li, N. Purpose based access control for privacy protection in relational database systems. *The VLDB Journal 17*, 4 (2008), 603–619.

[10] Capizzi, R., Longo, A., Venkatakrishnan, V. N., and Sistla, A. P. Preventing information leaks through shadow executions. In *Proceedings of the 2008 Annual Computer Security Applications Conference* (2008), IEEE Computer Society, pp. 322–331.

[11] Clark, D., and Hunt, S. Non-interference for deterministic interactive programs. In *Formal Aspects in Security and Trust* (2009), P. Degano, J. Guttman, and F. Martinelli, Eds., Springer-Verlag, pp. 50–66.

[12] Devriese, D., and Piessens, F. Noninterference through secure multi-execution. In *Proceedings of the 2010 IEEE Symposium on Security and Privacy* (2010), pp. 109–124.

[13] Enamul Kabir, M., Wang, H., and Bertino, E. A conditional purpose-based access control model with dynamic roles. *Expert Syst. Appl. 38* (2011), 1482–1489.

[14] FairWarning. Privacy breach detection for healthcare. White Paper, 2010.

[15] Goguen, J. A., and Meseguer, J. Security policies and security models. In *Proceedings of the IEEE Symposium on Security and Privacy* (1982), pp. 11–20.

[16] Hayati, K., and Abadi, M. Language-based enforcement of privacy policies. In *PET 2004: Workshop on Privacy Enhancing Technologies* (2005), Springer-Verlag, pp. 302–313.

[17] Jafari, M., Fong, P. W., Safavi-Naini, R., Barker, K., and Sheppard, N. P. Towards defining semantic foundations for purpose-based privacy policies. In *Proceedings of the first ACM conference on Data and application security and privacy* (2011), pp. 213–224.

[18] Jafari, M., Safavi-Naini, R., and Sheppard, N. P. Enforcing purpose of use via workflows. In *WPES '09: Proceedings of the 8th ACM Workshop on Privacy in the Electronic Society* (2009), pp. 113–116.

[19] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. Planning and acting in partially observable stochastic domains. *Artif. Intell. 101* (1998), 99–134.

[20] Kurniawati, H., Hsu, D., and Lee, W. S. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In *Proc. Robotics: Science and Systems* (2008).

[21] Madani, O. *Complexity Results for Infinite-Horizon Markov Decision Processes*. PhD thesis, University of Washington, 2000.

[22] Massacci, F., Mylopoulos, J., and Zannone, N. Hierarchical Hippocratic databases with minimal disclosure for virtual organizations. *The VLDB Journal 15*, 4 (2006), 370–387.

[23] Mayer, J. R., and Mitchell, J. C. Third-party web tracking: Policy and technology. In *IEEE Symposium on Security and Privacy* (2012), pp. 413–427.

[24] McCamant, S., and Ernst, M. D. A simulation-based proof technique for dynamic information flow. In *Proceedings of the 2007 Workshop on Programming Languages and Analysis for Security* (2007), ACM, pp. 41–46.

[25] McCullough, D. Noninterference and the composability of security properties. In *IEEE Symposium on Security and Privacy* (1988), pp. 177–186.

[26] McLean, J. A general theory of composition for trace sets closed under selective interleaving functions. In *Proceedings of the 1994 IEEE Symposium on Security and Privacy* (1994), p. 79.

[27] Monahan, G. E. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science 28*, 1 (1982), 1–16.

[28] Newsome, J., and Song, D. X. Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software. In *Proceedings of the Network and Distributed System Security Symposium* (2005), The Internet Society.

[29] Ni, Q., Bertino, E., Lobo, J., Brodie, C., Karat, C.-M., Karat, J., and Trombetta, A. Privacy-aware role-based access control. *ACM Trans. Inf. Syst. Secur. 13* (2010), 24:1–24:31.

[30] Office for Civil Rights. Summary of the HIPAA privacy rule. OCR Privacy Brief, U.S. Department of Health and Human Services, 2003.

[31] Papadimitriou, C., and Tsitsiklis, J. N. The complexity of Markov decision processes. *Math. Oper. Res. 12* (1987), 441–450.

[32] Peng, H., Gu, J., and Ye, X. Dynamic purpose-based access control. In *International Symposium on Parallel and Distributed Processing with Applications* (2008), IEEE Computer Society, pp. 695–700.

[33] Poupart, P., Kim, K.-E., and Kim, D. Closing the gap: Improved bounds on optimal POMDP solutions. In *Proceedings of the International Conference on Automated Planning and Scheduling* (2011), F. Bacchus, C. Domshlak, S. Edelkamp, and M. Helmert, Eds., AAAI.

[34] Ramírez, M., and Geffner, H. Plan recognition as planning. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence* (2009), C. Boutilier, Ed., pp. 1778–1783.

[35] Ramírez, M., and Geffner, H. Goal recognition over POMDPs: Inferring the intention of a POMDP agent. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence* (2011), T. Walsh, Ed., IJCAI/AAAI, pp. 2009–2014.

[36] Rummery, G. A., and Niranjan, M. On-line Q-learning using connectionist systems. Tech. Rep. CUEF/F-INFENG/TR 166, Cambridge University Engineering Department, 1994.

[37] Sabelfeld, A., and Myers, A. C. Language-based information-flow security. *IEEE Journal on Selected Areas in Communications 21*, 1 (2003), 5–19.

[38] Schmidt, C., Sridharan, N., and Goodson, J. The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence 11*, 1-2 (1978), 45 – 83.

[39] Smith, T., and Simmons, R. Point-based POMDP algorithms: Improved analysis and implementation. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (July 2005).

[40] Sondik, E. J. *The optimal control of partially observable Markov processes*. PhD thesis, Stanford University, 1971.

[41] Taylor, R. *Action and Purpose*. Prentice-Hall, 1966.

[42] Tschantz, M. C., Datta, A., and Wing, J. M. Formalizing and enforcing purpose restrictions in privacy policies. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy* (2012), pp. 176–190.

[43] Tschantz, M. C., and Wing, J. M. Extracting conditional confidentiality policies. In *Proceedings of the Sixth IEEE International Conferences on Software Engineering and Formal Methods* (2008).

[44] Vachharajani, N., Bridges, M. J., Chang, J., Rangan, R., Ottoni, G., Blome, J. A., Reis, G. A., Vachharajani, M., and August, D. I. RIFLE: An architectural framework for user-centric information-flow security. In *Proceedings of the 37th Annual IEEE/ACM International Symposium on Microarchitecture* (2004), pp. 243–254.

[45] Venkatakrishnan, V. N., Xu, W., DuVarney, D. C., and Sekar, R. Provably correct runtime enforcement of non-interference properties. In *Proceedings of the 8th International Conference on Information and Communications Security* (2006), Springer-Verlag, pp. 332–351.

[46] Wittbold, J. T., and Johnson, D. M. Information flow in nondeterministic systems. In *Proceedings of the IEEE Symposium on Security and Privacy* (1990), pp. 144–161.

[47] Yahoo! Privacy policy: Yahoo Mail, 2013.

[48] Yumerefendi, A. R., Mickle, B., and Cox, L. P. Tightlip: keeping applications from spilling the beans. In *Proceedings of the 4th USENIX Conference on Networked Systems Design and Implementation* (2007), pp. 12–12.

[49] Zhou, R., and Hansen, E. A. An improved grid-based approximation algorithm for POMDPs. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence* (2001), vol. 1, Morgan Kaufmann, pp. 707–714.

Appendix: Details of Example POMDP

Here we provide details about the network POMDP m_{ex}. Formally, the state space is $\mathcal{Q} = \{\mathsf{f}, \mathsf{m}\} \times \{\mathsf{f}, \mathsf{m}, \bot\} \times \{\mathsf{ad}_1, \mathsf{ad}_2, \mathsf{ad}_3, \varnothing\}$ with f, m, $\bot$, ad_i, and $\varnothing$ interpreted as in Section 3. For example, the state $\langle \mathsf{f}, \bot, \mathsf{ad}_2 \rangle$ indicates that the visitor is a female, the database does not record her gender, and the network has shown her ad_2. $\rho(\langle \mathsf{f}, \bot, \mathsf{ad}_2 \rangle) = 0$ since the visitor has already seen an ad.

The actions and states are related by the transition function $\tau : \mathcal{Q} \times \mathcal{A} \to \mathsf{Dist}(\mathcal{Q})$. $\tau(q, a)$ is a distribution over states such that for each state q', $\tau(q, a)(q')$ is the probability of the environment transition from state q to state q' by the network performing action a. While the network has uncertainty about the gender of the visitor, each action selected by network deterministically results in the next state. Thus, in this model, for all states q and actions a, the distribution $\tau(q, a)$ is always a *degenerate distribution* that assigns a probability of 1 to exactly one state. Let $\delta(q)$ denote the degenerate distribution assigning the probability of 1 to the state q. In our model, $\tau(\langle g, d, \varnothing \rangle, \mathsf{ad}_i) = \delta(\langle g, d, \mathsf{ad}_i \rangle)$ for all g in $\{\mathsf{f}, \mathsf{m}\}$, d in $\{\mathsf{f}, \mathsf{m}, \bot\}$, and i in $\{1, 2, 3\}$ reflecting that showing an advertisement does not change the visitor's gender or the network's database. $\tau(\langle g, d, \mathsf{ad}_i \rangle, \mathsf{ad}_j) = \delta(\langle g, d, \mathsf{ad}_i \rangle)$ since the network can show the visitor only one advertisement. $\tau(q, \mathsf{lookup}) = \delta(q)$ since looking up information in the database does not change the state of the environment.

The function $\nu : \mathcal{A} \times \mathcal{Q} \to \mathsf{Dist}(\mathcal{O})$ relates these observations to actions and states. Again we restrict our attention to degenerate distributions since our example contains uncertainty but not truly random processes. For each state $\langle g, d, \alpha \rangle$, the lookup action results in the observation $\langle d, \alpha \rangle$. For simplicity, we model actions of showing an advertisement as providing a similar observation. Thus, for all actions a, $\nu(a, \langle g, d, \alpha \rangle) = \delta(\langle d, \alpha \rangle)$. (Since the network only gets to show one advertisement and following actions do not affect its total reward, the observation made from showing an advertisement is of no consequence.)